

# Over- and Undereducation Among Swedish Immigrants

---

**Manuel Echeverría**

**2008-11-08**

**Supervisors: Inga Persson, Dan-Olof Rooth**

This thesis analyses the labour market mismatch among four different ethnic groups. The results reflect the importance of the transferability of human capital in the Swedish labour market. These results indicate that Nordic and Western immigrants have a greater transferability of human capital than Eastern and Southern Europeans and non-Western immigrants. Furthermore, Swedish schooling or being partly brought up in Sweden seems to be an important determinant of the labour market mismatch.

This thesis is a part of the project: *The Importance of Swedish-specific Human Capital on the Swedish Labour Market* (Betydelsen av Sverige-specifikt humankapital på svensk arbetsmarknad). Dan-Olof Rooth, University of Kalmar, project leader. Financial support from the Swedish Council for Working Life and Social Research (FAS) is gratefully acknowledged.

## Contents

<b>1 INTRODUCTION</b> .....	<b>5</b>
<b>2 PURPOSE</b> .....	<b>6</b>
<b>3 METHOD</b> .....	<b>8</b>
3.1 DIFFERENT KINDS OF MEASURES .....	8
3.2 THE MAGNITUDE OF OVEREDUCATION .....	12
3.3 SUMMARY.....	13
<b>4 THEORY</b> .....	<b>13</b>
4.1 THEORIES OF LABOUR MARKET MISMATCH AMONG IMMIGRANTS.....	14
4.2 HUMAN CAPITAL THEORY .....	14
4.3 TECHNOLOGICAL CHANGE THEORY.....	16
4.4 THE SCREENING HYPOTHESIS .....	16
4.5 ASSIGNMENT THEORY.....	17
4.6 DISCRIMINATION .....	17
4.7 SUMMARY.....	18
<b>5 THE DATA</b> .....	<b>19</b>
<b>6 THE INCIDENCE OF OVEREDUCATION</b> .....	<b>20</b>
6.1 YEARS SINCE MIGRATION AND LABOUR MARKET MISMATCH.....	20
6.2 THE IMPORTANCE OF AGE AT MIGRATION .....	22
6.3 YEARS SINCE MIGRATION AND MISMATCH AMONG FOUR IMMIGRANT GROUPS.....	25
6.4 THE MAGNITUDE OF MISMATCH .....	31
6.5 SUMMARY.....	31
<b>7 REGRESSION ANALYSIS OF THE MISMATCH AMONG IMMIGRANTS</b> .....	<b>32</b>
7.1 SPECIFICATION .....	32
7.2 RESULTS .....	34
7.2.1 <i>Overeducation and Undereducation among All Immigrants</i> .....	34
7.2.2 LABOUR MARKET MISMATCH AMONG FOUR ETHNIC GROUPS.....	36
7.2.3 <i>Models with Ethnicity and Age at Migration</i> .....	38
7.2.3.1 <i>Overeducation</i> .....	38
7.2.3.2 <i>Undereducation</i> .....	42
7.3 QUANDARIES.....	45

7.4 SUMMARY.....	46
<b>8 SUMMARY AND CONCLUSIONS.....</b>	<b>47</b>
<b>8 APPENDIX.....</b>	<b>51</b>
<b>REFERENCES.....</b>	<b>53</b>

TABLE 1 IMPLICATIONS OF LABOR MARKET THEORIES FOR IMMIGRANTS.....	17
TABLE 2 YEARS SINCE MIGRATION AND THE INCIDENCE OF OVEREDUCATION .....	21
TABLE 3 MISMATCH AND AGE AT MIGRATION (RM) .....	23
TABLE 4 MISMATCH AND AGE AT MIGRATION (JA) .....	23
TABLE 5 THE RM MISMATCH AND JOB REQUIREMENTS ACCORDING TO JA.....	24
TABLE 6 YEARS SINCE MIGRATION AND MISMATCH FOR NORDIC AND WESTERN IMMIGRANTS (RM) .....	26
TABLE 7 YEARS SINCE MIGRATION AND MISMATCH AMONG EASTERN AND SOUTHERN EUROPEAN AND NON- WESTERN IMMIGRANTS (RM) .....	27
TABLE 8 YEARS SINCE MIGRATION AND MISMATCH FOR NORDIC AND WESTERN IMMIGRANTS (JA) .....	28
TABLE 9 YEARS SINCE MIGRATION AND MISMATCH AMONG EASTERN AND SOUTHERN EUROPEAN AND NON- WESTERN IMMIGRANTS (JA) .....	28
TABLE 10 LABOUR OUTCOME AND REQUIREMENTS AMONG ETHNIC GROUPS .....	30
TABLE 11 MEAN YEARS OF UNDER- AND OVEREDUCATION AMONG THE ETHNIC GROUPS.....	31
TABLE 12 LABOUR MARKET MISMATCH AMONG ALL IMMIGRANTS.....	35
TABLE 13 LABOUR MARKET MISMATCH AMONG FOUR IMMIGRANT GROUPS.....	37
TABLE 14 LPM FOR OVEREDUCATION .....	39
TABLE 15 PROBABILITY DIFFERENTIALS FOR OVEREDUCATION.....	41
TABLE 16 AGE OF MIGRATION DIFFERENTIALS.....	42
TABLE 17 LPM FOR UNDEREDUCATION .....	43
TABLE 18 PROBABILITY DIFFERENTIALS FOR UNDEREDUCATION.....	44
TABLE 19 AGE AT MIGRATION DIFFERENTIALS.....	45
A. 1 THE SIGNIFICANCE OF AGE AT MIGRATION AND ETHNICITY.....	51
A. 2 THE SIGNIFICANCE OF AGE AT MIGRATION AND ETHNICITY.....	52
FIGURE 2 HUMAN CAPITAL FUNCTION .....	15

## **1 Introduction**

The aim of this study is to analyze the incidence of mismatch as indicated by over- or undereducation among four immigrant groups with focus on possible differences between them. These groups are Nordic, Western, Eastern and Southern European, and Non-Western immigrants, where the group Western contains immigrants from Western Europe, North America and Oceania. Labour market theories will be presented as candidates for an explanation of these differences. Immigrants share the special feature of moving from a foreign to the domestic, host country labour market. Taking this into account using the conventional labour market theories, relevant predictions can be made for the special immigrant case.

As a first step, the incidence of mismatch will be presented with descriptive statistics. In this way the specific features of different immigrant groups, defined after ethnicity, will appear. Theory suggests that a set of factors or relations could be important to the incidence of mismatch in general and in the immigrant case. New subsets will therefore be constructed from the basic ethnic groups in a way which relates to theory. By doing this, a first glance or hint of the impact from the relevant factors suggested by theory can be attained.

In a second step, the impact from different factors will be analyzed in a slightly more rigorous way. This is done through a linear probability model that assigns each group a probability of being overeducated (undereducated) given other factors of importance. Following the subsets outlined in the first part of the analysis with descriptive statistics, factors that affect the probability of being overeducated will be estimated. The patterns from this study will in turn be discussed in the light of the implications from the labour market theories in an immigrant context.

## 2 Purpose

The purpose of this thesis is to study the labour market mismatch as indicated by over- and undereducation among immigrant men. In order to do this, we start by investigating the theoretical reasons for the mismatch. The first intermediate aim is to establish an interpretation of over- and undereducation so that the empirical evidence can be analyzed in the light of theory. The first questions to be answered are:

*How can over- and undereducation be interpreted and what are the predictions for the immigrants' labour market mismatch?*

A related issue is to investigate how well immigrants do in the labour market by considering a link between labour market mismatch and labour market success or achievement. The motivation of this interpretation is based on the ORU framework which is briefly discussed in section three, and the *Search and Match theory* discussed in section four. This discussion is basically a comment to the ORU framework.

The measurements of labour market mismatch are discussed quite extensively because previous research shows that the results could be sensitive to the choice of measure. The aim of the methodological discussion is to pinpoint potential problems that could bias the results and in that way problemize the conclusions drawn from the statistical analysis. This is done in practice by combining different components of the various measures in the analysis.

The main issue of this thesis is to study the incidence of under- and overeducation among different immigrant groups, with special attention on the differences between 'Westerners' and immigrants from non- Western countries. Labour market theory suggests that other factors than ethnicity matters; the question to be answered is then:

*Is ethnicity an important determinant for the incidence of labour market mismatch or are other factors (suggested by theory) more important?*

This question can be reduced to:

*How can the observed mismatch among immigrants be explained?*

The objective in this respect is to draw conclusions about which theories are supported by evidence and in this way at least give some hints about how the mismatch is not to be explained.

### 3 Method

A worker is categorized as overeducated if she has an educational level above the required level for the job. The opposite is true for an undereducated worker. The categorization of correctly matched or mismatched workers depends on how the labour market mismatch is measured and how the boundaries for the categories are defined. It is important to note that in some few cases the discussion is more relevant when overeducation is used as an independent variable. In this thesis, under- and overeducation are dependent variables.

#### 3.1 Different Kinds of Measures

There are several ways of measuring the mismatch between the attained level and the *required* level of education. Basically a worker can be seen as correctly educated, under- or over-educated. Hartog (2000) describes three distinct ways to measure the required level of education; *job analysis* (JA), *worker self- assessment* (WA) and *realized matches* (RM). The JA is an evaluation of the requirements made by a professional job analyst, in the case of WA, the workers themselves are asked to evaluate the required level of schooling, their perceived over-education or any related question, e.g. the required level to do the job. The third approach is to measure and compute the actual mean or mode level of education for the different jobs. The RM and JA measures are used in this thesis.

The two first approaches (JA and WA) are conceptually closer to the notion of required education and could be linked to demand curve parameters, whereas the RM approach rather measures the actual *market outcome*. In spite of this, all three methods have their advantages and disadvantages.

This discussion could be summarized by looking at the way of measurement in a more general way by expressing it as an equation as it is done by Verhaest and Omey (2006).

$$mm_i = e_i - r_i$$

The educational mismatch ( $mm_i$ ) is defined by the difference between the individual's attained educational level ( $e_i$ ) and the required level ( $r_i$ ) to do the job. If this difference is positive, then the individual is overeducated. If the difference is negative then the individual is undereducated. Boundaries around the point  $r_i$  could then be used to give it a dimension;

otherwise every observation would be classified as a mismatch. This range  $|e_i - r_i|$  is called  $A$ , and is the minimum distance from  $r_i$  required for an individual to be classified as a mismatch. If  $e_i^m$  and  $r_i^m$  are the corresponding measurements for  $e$  and  $r$ , then the measurement of the mismatch incidence  $MM_i$  is defined by<sup>1</sup>:

$$\left\{ \begin{array}{l} MM_i = 1 \Leftrightarrow (e_i^m - r_i^m) \geq A \\ MM_i = 0 \Leftrightarrow |e_i^m - r_i^m| < A \\ MM_i = -1 \Leftrightarrow (e_i^m - r_i^m) \leq -A \end{array} \right. \Leftrightarrow \left\{ \begin{array}{l} MM_i = 1 \Leftrightarrow e_i^m \geq r_i^m + A \\ MM_i = 0 \Leftrightarrow r_i^m - A < e_i^m < r_i^m + A \\ MM_i = -1 \Leftrightarrow e_i^m \leq r_i^m - A \end{array} \right.$$

Where 1 denotes overeducation, 0 correctly matched and -1 undereducation. In this manner it can be clearly seen that the measured incidence is dependent on three factors, namely  $e$ ,  $r$  and  $A$ . (Verhaest and Omey 2006, p.785f)

As shown above, the natural way of defining over- and undereducation is by the upward and downward deviation from the required level. One approach is to define required education, as education within one standard deviation from the job requirements for an occupation. Over- and undereducation then becomes the actual levels of education that are more than one standard deviation above or below the *required* one for the occupation (measured by RM).

When measured with the RM and standard deviation approach, the incidence of over- and undereducation is commonly symmetrical and about 10-15 %. This is close to the tails beyond one standard deviation in a normal distribution. (Hartog 2000, p. 131-133)

The RM method described above does not capture minor differences between attained and mean education. Furthermore it implies a symmetry which is seldom found in practice; therefore it is likely to induce biased estimations. Some authors prefer the use of the modal over the mean; this is thought to lessen the impact of outliers and technological change.

The importance of formal education may vary among different occupations; some may have strict requirements while the need of formal schooling is less accentuated in others. One way of measuring this could be the dispersion of formal education within a given occupation. (Sloane 2002, p.8f)

---

<sup>1</sup> In the original paper the third equation in the system was, probably due to a typo, an inequality with  $A$  on the RHS.  $-A$  gives a more intuitive interpretation of the threshold.

The *realized matches* or statistical approach has been criticized for the arbitrary choice of threshold levels for under- and overeducation. Higher rates of overeducation in a given occupation will, *ceteris paribus*, yield a higher mean with higher corresponding mismatch-thresholds. This in turn implies an underestimation of the true overeducation level. One maybe more obvious explanation is the fact that it is only the statistical approach to measurement that has a threshold of one standard deviation for overeducation. (McGuinness 2006, p. 397, 397) These features are probably reflected in the significantly smaller reported incidence of overeducation compared to the other measurements, e.g. in a recent study, around 8 % were overeducated with the RM approach whereas the corresponding proportion was 51% using JA data. At the same time, the incidence of undereducation was higher compared to the other types of measurement, (Verhaest and Omey 2006, s. 791, see also McGuinness 2006 p.407f), something that could be attributed to the symmetry feature. These types of problems are linked to factor A and  $r$ .

In general these three ways of measurement fail to capture the underlying differences in skills and ability between individuals with the same level of education. Substitutability between informal skills or ability and formal education implies unobserved differences in actual skills and observed skills identified as formal education. This discrepancy means that a seemingly overeducated individual in fact could be correctly educated or undereducated. This phenomenon is illustrated graphically in figure 1. Moreover, even if the worker is correctly matched with respect to the level of education, this could still be a case of mismatch if the *type* of education is incorrect. (Sloane 2002, p.7, 9, 11f) These kinds of problems can be attributed to the factors  $e$  and  $r$ .

Similar problems stem from the measurement of  $r$  which is the most discussed factor and therefore presented in more detail. It can be thought of being formulated as:

$$r_i^m = r_i + a + X_i b + \mu_i$$

According to this equation, there are three channels of measurement errors.

The first channel is the random error  $\mu_i$  which in turn has two main error sources which are created when the data is classified. When data is grouped to a limited number of qualification levels, a given educational category could in fact contain a range of skill levels. Data collection and processing is the second source; these sources imply misclassification of jobs. This is not a problem if the error is not too large to displace the estimation beyond the

threshold value  $A$ . (Verhaest and Omey 2006, p.788) The choice of the threshold  $A$  can affect the incidence of mismatch. Verhaest and Omey show that a strong positive trend exists between the overeducation rate and the number of educational categories in the measure.<sup>2</sup> (Ibid. p. 792)

The second and most discussed channel (a) is the systematic estimation error of the required level of education. Technological change could be a source of systematic bias in the case of JA. If this change is skill biased the result is a systematic underestimation of the required level. In the case of RM, the business cycles could also be a factor if employers adapt the requirements to the cycle. A related source of bias is the existence of an oversupply of educated workers, resulting in an overestimation of the required level. In addition, the RM approach does not account for skills other than formal schooling; this tends to underestimate the required level. If the RM measure is interpreted literally, an implicit assumption of flexible markets is made. This implies that the required level equals the mean or modal level of the occupation. (Verhaest and Omey 2006, p.789)

The other systematic bias (b) is channeled through individual characteristics  $X_i$  and is potentially very problematic when comparison of the overeducation outcome is made between workers or jobs with different characteristics. The impact of technological change may be different for different occupations and consequently affect the bias of the JA measure differently among these jobs. Imbalances of the supply and demand for educated workers may vary between occupations; a shortage may induce a reduction of the required level to get the job. It follows that the real incidence of overeducation will be biased with the RM approach. In addition the RM method is flawed when it comes to comparison of the mismatch by different educational levels.

Different methods generate different results, both in terms of the magnitude of the mismatch rate and in terms of consistency in the classification of individuals. In other words, there can be a low correlation between different measures. (Bourdet & Persson 2008, s.6) More importantly, the conclusions typically drawn from this kind of analysis are in spite of these inconsistencies not necessarily sensitive to the choice of method. (Chiswick 2007, p.11) These differences are connected to the different ways of measuring  $e$ ,  $r$  and  $A$  as discussed above.

---

<sup>2</sup> This was computed for two types of WA measures.

As the name hints, the *realized matches* (RM) method could be interpreted as the actual labour market outcome. The explanation of this particular outcome could in turn be derived from different theoretical frameworks with varying predictions of the labour market equilibrium.

### 3.2 The Magnitude of Overeducation

The extent of over- and undereducation has so far been seen as a discrete outcome. But the magnitude of over- and undereducation has the potential to further illuminate the mismatch conditions. A year of overeducation is here defined as one year over what is considered as the required education, i.e.  $r$  plus  $A$ . This is in principle the measure used in the ORU<sup>3</sup> approach to model the returns to years of educational mismatch. (See e.g. Voon & Miller 2005, p.26) If an individual is several years, maybe several standard deviations above the normal requirements, then this individual can be seen as more overeducated than an individual who is a couple of months over the normal level.

Another way of measuring this using the components of the JA measure, is defining an individual who is more than two categories above or under the required level as more over- or undereducated than a corresponding individual who deviates from the required level by exactly one category.

It is therefore of interest to study if certain groups of individuals have a more severe mismatch than others. This relates to the weakness of the measures discussed above and gives an additional dimension to the discussion of mismatch. Two groups that have the same distributions of correctly educated, undereducated and overeducated may still differ in the mean numbers of years of overeducation and undereducation.

This is interesting for the purpose of a discussion which relates the mismatch to what can be described as under- and overachievement in the labour market or in other words, how successful the individual is in the labour market. One motivation of this interpretation is based on the different returns to over- and undereducation. A year of overeducation usually gives a different return than a year of undereducation in an ORU framework. So a measure of the years of mismatch, which are associated to different payoffs, is a way to account for some important determinants of income.

---

<sup>3</sup> The typical result is that the return to one year of overeducation is lower than the return to one year of required education. Moreover an undereducated worker has a lower return but will still earn more than in a job which fits his attained level of education. (Bourdet & Persson 2008, p.12)

### 3.3 Summary

The mismatch measure can be described as being constructed of three different parts, each of which can be made in different ways. Different ways of constructing the dependent or outcome variable, lead to different problems which can be attributed to the components  $A$ ,  $e$  and  $r$ .

Two measurements are used in this thesis, one which is categorized as a statistical, *realized matches* measure (RM) and the other which essentially is a *Job Analysis* measure (JA). The RM type measure can be interpreted as the actual labour market outcome indicator of mismatch, while the JA type is arguably closer to the notion of *required education*.

To give a more complete picture of the mismatch, the magnitude of over- and undereducation is also considered. This measure can distinguish which group is the *most* under- or overeducated by measuring the number of years of mismatch.

## 4 Theory

A set of labour market theories with focus on their implications for immigrants are presented in this section. The main point is the transferability of immigrant human capital and its impact on overeducation among immigrants. In addition a variant of discrimination in the context of overeducation is discussed. In this thesis discriminatory practices are simply thought to be an extra difficulty to transfer the human capital in addition to what could be explained by the standard theories. Discrimination is then in a way treated as a residual. This part of the thesis helps us to pick candidates of explanatory variables which must be accounted for when analyzing the determinants of the mismatch. The theoretical discussion in this section is primarily concerned with the measurements. The analysis carried out here aims to achieve a greater understanding of what the measurements capture and which concepts are associated to these.

Chiswick (2007) describes five theories which could explain the overeducation phenomenon: search and match theory, human capital theory, assignment theory, technological change theory and the screening hypothesis.

#### **4.1 Theories of Labour Market Mismatch Among Immigrants**

*The Search and Match Theory* suggests that overeducation is a transitory phase connected to labour market entry and due to imperfect information. Overeducated labour market entrants could over time be expected to obtain jobs which better correspond to their level of education. The overeducation mismatch is expected to drop with labour market experience as the information for the supply and demand sides improves.

Moreover, there is a hypothesis about the impact of the search and match process on undereducation. Assuming wealth-maximizing behavior and that jobs with higher requirements give higher wealth, a worker will only change job if the educational requirements are lower than the worker's attained level. In this case undereducation will tend to rise with the searching and matching process over time. This is the second motivation to consider some undereducated individuals as more 'successful' (the first one is found in the ORU- framework which typically concludes that undereducated individuals have higher returns than they would have if they had an occupation which matched the requirements.).

Information problems connected to labour market entrance are supposedly common among immigrants entering a foreign market. Immigrants from distant and very different labor markets should be affected by their own and the employers' limited information about jobs and qualifications. (Chiswick 2007, p.5f)

#### **4.2 Human Capital Theory**

The core assumption of the *human capital theory* is that workers are paid their marginal product. Firms are accordingly supposed to adapt their production in response to any change in the relative supply of labour to fully utilize the workers' human capital stock. The workers accumulate skills through formal education and on the job training. Because their skills are fully utilized, overeducation should be nonexistent and regarded as inconsistent with the theory. (McGuinness 2006, p.389f)

There are several ways to make the observed mismatch consistent with human capital theory. Overeducation could be seen as a temporary short run disequilibrium when the firm's other factors of production are constant and the firm is unable to adapt the production process. In this time horizon, workers themselves could also have problem to find a suitable job.

Adjustments on the supply and demand side should be possible to make in the long run and the human capital theory holds. (Ibid, p.390f).

Another way to explain overeducation is the decomposition of human capital accumulation into formal education and on the job training. Formal and informal human capital are substitutable; more of the former could compensate for lack of the other. If less formal human capital is not accounted for, then seemingly overeducated workers could in fact be lacking informal human capital. An undereducated individual could conversely have an abundance of unobserved skills.

**Figure 1 Human Capital Function**

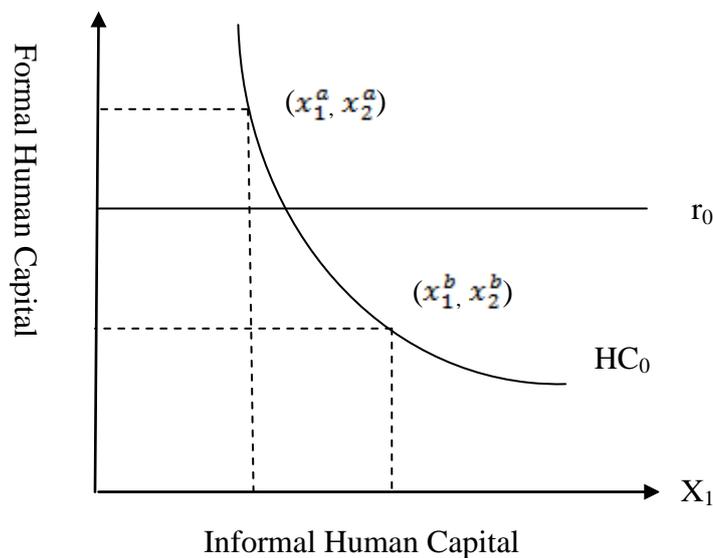


Figure 2 illustrates the tradeoff between formal and informal human capital and the mean required level of education  $r$  in an occupation. For a given level of total human capital, there exist an arbitrarily large number of pairs  $(X_1, X_2)$  which are consistent with the level. Total human capital is in this case thought to be achieved by different combinations of  $X_1$  and  $X_2$ .  $(x_1^a, x_2^a)$  illustrates a seemingly overeducated individual and  $(x_1^b, x_2^b)$  the opposite case. This is a consequence of perceiving (or measuring) human capital by only  $X_2$ , this is an omitted variable problem. (McGuinness 2006, p.390)

Ability works in a similar manner. Since more able individuals are more productive and ability is more difficult to observe, a more able individual has a higher skill level for any pair of formal and informal human capital than a less able individual.<sup>4</sup>

It follows that a more able individual with a higher skill level can be seen as undereducated even if informal human capital is accounted for.

Transferability of human capital can in this context be thought of as a factor which devaluates the immigrant's formal and informal human capital and hence her total human capital level.<sup>5</sup> (See e.g. Orcutt et al., 2008 for an example of the use of a transferability coefficient)

Note that all these examples only address the issue of what is possible to observe, they address the issue of measurement. It can be tempting to relate these observations to a more theoretical discussion but this is beyond the scope of this thesis.<sup>6</sup>

### 4.3 Technological Change Theory

*Technological Change Theory* relates the observed mismatch to a temporary labour market disequilibrium induced by technological development. Increased formal schooling is viewed as a response from the labour force to changes in skill requirements due to technological progress. The share of the workforce engaged in production will become less educated than the ones who acquired the new skills. The replacement of the less educated workforce will be delayed due to adjustment costs. In time there will be an adaptation as companies will hire the new breed of more educated workers. The mismatch is therefore seen as a disequilibrium phenomenon. Relatively undereducated workers are the older ones who exist in a transition period. The overeducated workers are the more recent labour market entrants.

### 4.4 The Screening Hypothesis

*The Screening Hypothesis* states that workers signal their unobserved ability through schooling. There is a small initial mismatch which is accentuated as individuals with high unobserved ability are promoted and individuals with low unobserved ability are demoted.

---

<sup>4</sup> E.g. If  $y = Af(x_1, x_2) = Ax_1^\alpha x_2^{1-\alpha}$  then  $A^*f(x_1, x_2) > Af(x_1, x_2) \forall (x_1, x_2)$  if  $A^* > A$  assuming positive values and  $A^*, A > 1$ . And  $A$  stands for ability. (see e.g. Jones 2002, p.45)

<sup>5</sup> E.g.  $f(\tau_1 x_1 \tau_2, x_2) = \tau f(x_1, x_2) \leq f(x_1, x_2) \forall (x_1, x_2)$  if  $\tau_1 = \tau_2$  and  $\tau \in (0, 1)$  (see e.g. Varian 1992, p.17)

<sup>6</sup> The static nature of the analysis may in several cases be erroneous in a dynamical theoretical framework where adjustments on the supply and demand side are crucial.

Assuming risk-averse employers that are unsure of the immigrant's schooling signals, then the level of overeducation is hypothesized as initially high and decreasing with labour market experience in the destination country.

This perspective can neither explain initial undereducation or the variation in the incidence of undereducation over time in the destination country.

#### 4.5 Assignment theory

This set of theories emphasizes the importance of job characteristics as well as individual characteristics for the labour market outcome. In equilibrium there can be a discrepancy between required and attained education of the individuals but there are no specific immigrant implications according to Chiswick (2007).

Table 1 summarizes the implications for immigrant mismatch of the five theories

**Table 1 Implications of Labor Market Theories for Immigrants**

Theory	Undereducation	Overeducation
<b>(i) Search and Match Theory</b>	<ul style="list-style-type: none"> <li>No implications (Could rise with duration of residence)</li> </ul>	<ul style="list-style-type: none"> <li>More prevalent for immigrants from countries with labor markets and institutions distant from the destination country</li> <li>Declines with duration of residence</li> </ul>
<b>(ii) Human Capital</b>	<ul style="list-style-type: none"> <li>More prevalent for immigrants than for the native born due to self selection in migration and ability/motivation substituting for human capital</li> <li>No change with duration of residence</li> </ul>	<ul style="list-style-type: none"> <li>More prevalent among immigrants than for native born due to less-than-perfect international transferability of human capital</li> <li>Declines with duration of residence</li> <li>More prevalent for immigrants from countries with labor markets and institutions distant from the destination country</li> <li>Declines with duration of residence</li> </ul>
<b>(iii) Screening</b>	<ul style="list-style-type: none"> <li>No implications</li> </ul>	<ul style="list-style-type: none"> <li>Declines with duration of residence</li> </ul>
<b>(iv) Technological change</b>	<ul style="list-style-type: none"> <li>No implications</li> <li>No change with duration of residence</li> </ul>	<ul style="list-style-type: none"> <li>More prevalent for immigrants from less developed countries</li> </ul>
<b>(v) Assignment Theory</b>	<ul style="list-style-type: none"> <li>No implications</li> </ul>	<ul style="list-style-type: none"> <li>No implications</li> </ul>

(Chiswick 2007, p.9)

#### 4.6 Discrimination

Overeducation may also be more common among ethnic minorities due to hiring discrimination. (Sloane 2002, p.28) If discriminatory practices lead to the rejection or devaluation of immigrant merits, then immigrants should be overeducated to a greater extent. Discrimination could be a factor to consider and can arguably be expected to be more prevalent among immigrants from outside Europe. One way of thinking of the impact of

discrimination on overeducation, is that it is an additional, arbitrary obstacle to the transferability of immigrant human capital. Immigrants with Swedish education would then still tend to be more overeducated than natives, non- Western immigrants more overeducated than European immigrants. This can be further illustrated by assuming the human capital function above and introducing a discrimination coefficient that lowers the transferability of human capital due to discrimination in a framework of statistical discrimination. The implication is that even if two immigrants have the same level and transferability of human capital, but one is discriminated, then the discriminated individual's level of human capital will still be *perceived* as lower by the employer.<sup>7</sup>

Note that this discussion relates to the previous methodological section and is carried out in a semi-formal way. A rigorous analysis to deduct if these statements can be carried over to the underlying theories is beyond the scope of this thesis. The motivation of this semi formal model is only to problemize what theory proposes and what is possible to observe on the other hand.

Some of the implications from these set of theories overlap and some are complementary and could be used to define relevant categories of immigrants.

#### 4.7 Summary

There are three theories with implications on undereducation. The modified *Search and Match Theory* states that undereducation could rise with duration of residence. The human capital theory implies, on the other hand, that undereducation is more common among immigrants, but this does not change with duration of residence. Four of the theories have implications for overeducation, predicting that overeducation is more common among immigrants and three of these predict that overeducation declines with duration of residence.

Hence, there is a strong case for overeducation to decline with years since migration and undereducation could in some cases be expected to rise over time.

Labour market success or achievement can be linked to over- and undereducation through the ORU-framework and the *Search and Match Theory*.

The incidence of mismatch is complicated by the fact that ability and discrimination is unobserved.

---

<sup>7</sup> Eg.  $\tau_d f(x_1, x_2) < \tau f(x_1, x_2)$  where  $\tau_d < \tau$  is the discrimination coefficient and  $\tau_d, \tau \in (0,1)$

## 5 The Data

This section presents the data and the restrictions made in this study with special focus on how variables are constructed. The empirical analysis is based on a data set constructed by integrating three registers at Statistics Sweden (SCB), which identifies individuals by their social security numbers. The first register used is the Register of the Total Population (RTB), which contains information about some demographic and socioeconomic variables. Information on individual's level of education and occupation is taken from the Educational and Occupational register. The total data set contains information on about 323,718 immigrants that were between 16 and 64 years old in 2003.

The cross sectional data set is restricted to immigrant men of 25-55 years, a reasonable interval in a labour market context. This set has 290,592 observations from 2003. This set is restricted to the cases with information on schooling, year of migration and occupation. These restrictions may bias the results. The missing information could unfortunately be from interesting cases. In the case of the choice of gender on the other hand, selecting immigrant women may in fact give rise to even more selection bias because working women from certain countries may not be very representative. These restrictions leaves us with a set of 179,893 cases containing information enough to compute *years since migration*, *potential experience*<sup>8</sup> and the RM and JA measures. The RM measure is computed from a larger data set consisting of 704,375 Swedish men of 35-45 years of age which is reduced to 594072 when selecting individuals with the relevant information for computing the measure. The labour market mismatch among immigrants is therefore measured relative to the native male population.

Attained education is computed from SUN2000 that is adjusted to international standards (ISCED97). Information on the occupations is computed from SSK3. The requirements for the jobs, used for the JA method, are described in *Standard för svensk yrkesklassifiering* (2001, p.11) and are divided in to five skill levels<sup>9</sup>. A similar way of computing the JA measure is presented in a paper by Johansson & Katz (2007).

The RM measure is computed by taking the means and standard deviations of education from SUN2000 for each occupation classified after SSK3.

---

<sup>8</sup> Exp=age-years of schooling-7

<sup>9</sup> 1) Normally small or no requirements 2) Normally secondary school 3) Normally secondary school with extension or shorter university education 4) Normally a longer university education (3-4 years or more) and completed degree. There exists moreover occupations that have undefined requirements, these jobs are restricted to military or leadership related occupations.

The data is further restricted to immigrants that spent at least five years in Sweden, who do not belong to the occupations with undefined requirements and who do not have unidentified origin. These last restrictions leave us with a data set of 149,420 individuals.

## **6 The Incidence of Overeducation**

In this section, basic facts of the sample and the incidence of mismatch among immigrants are presented. The distribution of over-, under- and correctly educated is shown for immigrant groups with different characteristics. The incidence of mismatch is measured by the RM and JA methods discussed above.

### **6.1 Years Since Migration and Labour Market Mismatch**

To begin with it is of interest to look at how the incidence of under- and overeducation for all immigrants tends to change with respect to the time spent in Sweden since migration.<sup>10</sup> These tables also reflect cohort effects. The more educated younger cohorts might face a labour market with relatively few occupations with high requirements. Older cohorts could on the other hand have entered the labour market when occupations with higher requirements were more available. Lower rates of overeducation among workers with many years of experience or years since migration may partly be cohort effects. (McGuinness 2006, p.411)

The overall rate of overeducation is about 19% with the RM method and around 20% using JA. In the case of undereducation, the share is about 21% using the RM method and nearly 27 % with the JA.

---

<sup>10</sup> Similar tables are presented by Chiswick (2007) and Voon & Miller (2005).

**Table 2 Years Since Migration And The Incidence of Overeducation**

Years Since Migration		Incidence of Mismatch (RM)				Incidence of Mismatch (JA)			
		Normal	Undereducated	Overeducated	Total	Normal	Undereducated	Overeducated	Total
6-10	n	19464	5492	10159	35115	18748	6096	10271	35115
	(%)	<b>55.4%</b>	<b>15.6%</b>	<b>28.9%</b>	<b>100.0%</b>	<b>53.4%</b>	<b>17.4%</b>	<b>29.2%</b>	<b>100.0%</b>
11-15	n	18736	6751	7536	33023	16791	7905	8327	33023
	(%)	<b>56.7%</b>	<b>20.4%</b>	<b>22.8%</b>	<b>100.0%</b>	<b>50.8%</b>	<b>23.9%</b>	<b>25.2%</b>	<b>100.0%</b>
16-20	n	12778	3990	3501	20269	10966	5544	3759	20269
	(%)	<b>63.0%</b>	<b>19.7%</b>	<b>17.3%</b>	<b>100.0%</b>	<b>54.1%</b>	<b>27.4%</b>	<b>18.5%</b>	<b>100.0%</b>
21-25	n	11113	3733	2455	17301	9108	5350	2843	17301
	(%)	<b>64.2%</b>	<b>21.6%</b>	<b>14.2%</b>	<b>100.0%</b>	<b>52.6%</b>	<b>30.9%</b>	<b>16.4%</b>	<b>100.0%</b>
26-30	n	10682	4130	1975	16787	8638	5981	2168	16787
	(%)	<b>63.6%</b>	<b>24.6%</b>	<b>11.8%</b>	<b>100.0%</b>	<b>51.5%</b>	<b>35.6%</b>	<b>12.9%</b>	<b>100.0%</b>
31+	n	17586	6971	2368	26925	14291	10123	2511	26925
	(%)	<b>65.3%</b>	<b>25.9%</b>	<b>8.8%</b>	<b>100.0%</b>	<b>53.1%</b>	<b>37.6%</b>	<b>9.3%</b>	<b>100.0%</b>
Total	n	90359	31067	27994	149420	78542	40999	29879	149420
	(%)	<b>60.5%</b>	<b>20.8%</b>	<b>18.7%</b>	<b>100.0%</b>	<b>52.6%</b>	<b>27.4%</b>	<b>20.0%</b>	<b>100.0%</b>

With the RM method, the overall trend is that undereducation increases with time in Sweden, while overeducation decreases. This is consistent with the implications of labour market theories which were presented above. Undereducation increases from around 15% for immigrants with 6—10 years in Sweden to 25,6 % for the group who has been in Sweden for over 30 years, an increase of about 10 percentage points. Overeducation declines more sharply, from about 29 % to 9 %, a decline of approximately 20 percentage points.

These trends seem to be insensitive to the choice of outcome variable. The JA method gives similar results but with a greater magnitude of the mismatch. In addition, both the increase of undereducation and the decline of overeducation are now around 20 percentage points. This seems to some extent to be consistent with the previously discussed feature of the JA measure. As mentioned above, the studies with the JA method usually report a greater extent of mismatch, but the figures are not near the 50% overeducation rates which have been found in some cases. (Verhaest & Omey 2006, p.783) The magnitudes of overeducation are in fact strikingly similar for the two methods.

This gives support to the theoretical presumptions above. The increase of undereducation gives interestingly enough strong support to the *search and match* theory in relation to the somewhat weaker theoretical support. It is on the other hand worth noticing that the trend is not a monotonous increase because the share of undereducated falls somewhere between the

second and the third years since migration category using the RM measure. The tendency of overeducation to decline has a stronger theoretical base and this coincides with the sharper, theory consistent development with years since migration.

## **6.2 The Importance of Age at Migration**

One special circumstance which influences the transferability of human capital and is therefore important to highlight, is the age at migration. If the immigrant arrived at a low age e.g. before 16, then at least part of the compulsory education is Swedish. Moreover, the immigrant's upbringing could be expected to be at least somewhat more integrated, e.g. subtleties in the Swedish language and culture are arguably more accessible at a young age. An individual who is introduced to the Swedish society early on should have a better chance to adapt her human capital accumulation to the host country labour market conditions and her human capital stock should include a higher degree of transferability. The immigrant with a low age at migration would, in other words, be more similar to the native born than the one with an adult age at migration.

Another way of stating the importance of age at migration is to relate it more directly to theory. This circumstance influences the information available to the employer and employee concerning the workers' human capital; it improves the transferability of human capital and the signaling according to the screening hypothesis. With this in mind we could expect to observe a lower incidence of overeducation in this group.

There is an important problem when analyzing the age at migration with descriptive statistics, constructing groups will result in different selection problems. The individuals which belong to any of these groups will probably differ, not only regarding the distribution of labour market mismatch, but also in other important features which may affect the labour market mismatch. Immigrants with low age at migration have in general more years since migration, which is thought to influence the labour market outcome of interest. This issue is approached through a regression analysis in section seven. The motivation of the following descriptive study in this section is mainly to discuss the RM and JA measures and the relation between over- and undereducation and the notion of labour market success or achievement discussed in section three and four.

The importance of at least a partly Swedish schooling and upbringing can be illuminated by studying the incidence of mismatch for the groups with an age at migration greater or less

than 16. One natural hypothesis is that people who migrated to Sweden at a low age tend to be undereducated to a greater extent and, conversely, less overeducated.

**Table 3 Mismatch and Age at Migration (RM)**

		Age at Migration under 16*Education Crosstabulation			
		RM			
		normal	undereducated	overeducated	Total
Age of migration >16	n	61203	22991	22648	106842
	%	<b>57.3%</b>	<b>21.5%</b>	<b>21.2%</b>	<b>100.0%</b>
Age of migration <16	n	29156	8076	5346	42578
	%	<b>68.5%</b>	<b>19.0%</b>	<b>12.6%</b>	<b>100.0%</b>
Total	n	90359	31067	27994	149420
	%	<b>60.5%</b>	<b>20.8%</b>	<b>18.7%</b>	<b>100.0%</b>

This hypothesis seems to have support regarding overeducation but does not fit the data in the case of undereducation when using the RM method.

**Table 4 Mismatch and Age at Migration (JA)**

		Age at Migration under 16*Education Crosstabulation			
		JA			
		normal	undereducated	overeducated	Total
Age of migration >16	n	55177	27467	24198	106842
	%	<b>51.6%</b>	<b>25.7%</b>	<b>22.6%</b>	<b>100.0%</b>
Age of migration <16	n	23365	13532	5681	42578
	%	<b>54.9%</b>	<b>31.8%</b>	<b>13.3%</b>	<b>100.0%</b>
Total	n	78542	40999	29879	149420
	%	<b>52.6%</b>	<b>27.4%</b>	<b>20.0%</b>	<b>100.0%</b>

Assuming that the JA method is closer to the notion of required education and the RM method on the other hand is closer to the measurement of actual market outcome, then the difference between the measurements with respect to undereducation could at least partly be due to the construction of the measurements.

The JA method is in principle not affected by the distribution of schooling in a certain occupation. In view of this and the previously mentioned interpretation of the mismatch using JA, then the mismatch could be assumed to measure the actual discrepancy between the required level of schooling in a certain occupation and the attained level. The intuition behind the fact of the higher share of undereducated and the lower share of overeducated for the immigrants with a low age at migration, is that these immigrants are to a greater extent “overachievers”. These immigrants have, as previously mentioned, probably been better

integrated in society and the Swedish labour market; this feature could be reflected in these numbers.

The RM measure is on the other hand assumed to measure the actual labour market outcome, so the mismatch has a slightly different interpretation. These immigrants have a low education in relation to all the others in the same occupation. In this case the incidence of undereducation is smaller for immigrants with Swedish compulsory schooling but the difference in this case is on the other hand not very substantial (1.5 percentage points).

Even if immigrants raised in Sweden were more 'successful' in the labour market than those who migrated at an adult age, this may still not be captured by the RM measurement. A way of seeing this is to remember that an undereducated (RM measure) individual in a job without formal educational requirements is relatively undereducated compared to others in the same education but not in relation to the requirements of the job. This could affect many immigrants with compulsory education in jobs with low educational requirements in a Swedish context where most people have at least secondary education.

It is possible to address this problem by combining the RM measure of *labour market outcome* and the JA measure of job requirements. This is illustrated in table 5:

**Table 5 The RM Mismatch and Job Requirements According to JA**

			Age of migration under 16 * requirement Crosstabulation				
Mismatch (RM)			Requirement (JA)				Total
			No educational requirements	Secondary	Lower Tertiary	Higher Tertiary	
Normal	Age of migration >16	n	7902	36437	5722	11142	61203
		(%)	<b>12.9%</b>	<b>59.5%</b>	<b>9.3%</b>	<b>18.2%</b>	<b>100.0%</b>
	Age of migration <16	n	2291	17270	5297	4298	29156
		(%)	<b>7.9%</b>	<b>59.2%</b>	<b>18.2%</b>	<b>14.7%</b>	<b>100.0%</b>
	Total	n	10193	53707	11019	15440	90359
		(%)	<b>11.3%</b>	<b>59.4%</b>	<b>12.2%</b>	<b>17.1%</b>	<b>100.0%</b>
<b>Undereducated</b>	Age of migration >16	n	4371	15517	1153	1950	22991
		(%)	<b>19.0%</b>	<b>67.5%</b>	<b>5.0%</b>	<b>8.5%</b>	<b>100.0%</b>
	Age of migration <16	n	853	5393	829	1001	8076
		(%)	<b>10.6%</b>	<b>66.8%</b>	<b>10.3%</b>	<b>12.4%</b>	<b>100.0%</b>
	Total	n	5224	20910	1982	2951	31067
		(%)	<b>16.8%</b>	<b>67.3%</b>	<b>6.4%</b>	<b>9.5%</b>	<b>100.0%</b>
Overeducated	Age of migration >16	n	3044	14125	3148	2331	22648
		(%)	<b>13.4%</b>	<b>62.4%</b>	<b>13.9%</b>	<b>10.3%</b>	<b>100.0%</b>
	Age of migration <16	n	450	3197	1108	591	5346
		(%)	<b>8.4%</b>	<b>59.8%</b>	<b>20.7%</b>	<b>11.1%</b>	<b>100.0%</b>
	Total	n	3494	17322	4256	2922	27994
		(%)	<b>12.5%</b>	<b>61.9%</b>	<b>15.2%</b>	<b>10.4%</b>	<b>100.0%</b>

Most of the undereducated (RM method) are individuals in jobs which require secondary education (JA), i.e. these individuals have an education which is more than one standard deviation less than the mean education in occupations which are thought to require secondary schooling.

The share of individuals who are undereducated and in occupations with high educational requirements is greater for individuals who arrived to Sweden before 16. On the other hand, the share of undereducated in jobs with no educational requirements is much larger for individuals with an age at migration greater than 16. Linking higher educational requirements with high profile jobs gives a picture where immigrants who arrived to Sweden at a low age could be more 'successful' in the labour market as they often seem to be more undereducated in high profile occupations.

One explanation to the slightly higher share of undereducated among the group of immigrants who were older when they came to Sweden is that these men are more often undereducated in jobs with little or no requirements. These men have a low standard of education which is below secondary schooling with jobs which do not require formal schooling. These men meet the job requirements but have less schooling than others in the same Swedish occupation where secondary school is more common.

This illustrates the initial statement that the RM is closer to the actual labour market outcome and obscures in this case the relationship between the workers education and the requirements of the job.

### **6.3 Years Since Migration and Mismatch Among Four Immigrant Groups**

Table 6 and 7 make use of the RM measure and show how the labor market matching between attained and required education changes with years since migration, which is one potentially important candidate to explain differences in the mismatch outcome. The mismatch distribution for each of these groups of immigrants resembles the one observed for the sum of these groups. Undereducation is more, and overeducation less common among immigrant men who lived in Sweden longer.

**Table 6 Years Since Migration and Mismatch for Nordic and Western Immigrants (RM)**

Years Since Migration	Nordic				Western			
	Normal	Undereducated	Overeducated	Total	Normal	Undereducated	Overeducated	Total
6-10	936	322	453	1711	1310	303	908	2521
	<b>54.7%</b>	<b>18.8%</b>	<b>26.5%</b>	<b>100.0%</b>	<b>52.0%</b>	<b>12.0%</b>	<b>36.0%</b>	<b>100.0%</b>
11-15	1992	861	464	3317	1044	298	514	1856
	<b>60.1%</b>	<b>26.0%</b>	<b>14.0%</b>	<b>100.0%</b>	<b>56.2%</b>	<b>16.1%</b>	<b>27.7%</b>	<b>100.0%</b>
16-20	1407	609	282	2298	842	239	296	1377
	<b>61.2%</b>	<b>26.5%</b>	<b>12.3%</b>	<b>100.0%</b>	<b>61.1%</b>	<b>17.4%</b>	<b>21.5%</b>	<b>100.0%</b>
21-25	2261	1044	403	3708	835	256	212	1303
	<b>61.0%</b>	<b>28.2%</b>	<b>10.9%</b>	<b>100.0%</b>	<b>64.1%</b>	<b>19.6%</b>	<b>16.3%</b>	<b>100.0%</b>
26-30	3891	1972	503	6366	937	296	218	1451
	<b>61.1%</b>	<b>31.0%</b>	<b>7.9%</b>	<b>100.0%</b>	<b>64.6%</b>	<b>20.4%</b>	<b>15.0%</b>	<b>100.0%</b>
31+	11172	5022	1262	17456	1605	471	326	2402
	<b>64.0%</b>	<b>28.8%</b>	<b>7.2%</b>	<b>100.0%</b>	<b>66.8%</b>	<b>19.6%</b>	<b>13.6%</b>	<b>100.0%</b>
<b>Total</b>	21659	9830	3367	34856	6573	1863	2474	10910
	<b>62.1%</b>	<b>28.2%</b>	<b>9.7%</b>	<b>100.0%</b>	<b>60.2%</b>	<b>17.1%</b>	<b>22.7%</b>	<b>100.0%</b>

Immigrants from the Nordic countries have the highest overall rates of undereducation and the lowest rates of overeducation. The total difference between the ones who lived up to 10 years and those with more than 31 years in Sweden is about 10 percentage points, in the case of undereducation and 20 percentage points for overeducation. These immigrants have a pronounced change in the mismatch between the first two categories of years since migration (7 percentage points and 12 percentage points change in under and overeducation). Worth noting is that the typical Nordic immigrant has lived in Sweden more than a decade, most of them have lived in Sweden for more than three decades, the mean value is around 29 years. It is important to bear in mind that these figures partly reflect cohort effects. The composition of immigrants from different waves of migration to Sweden differ through out the modern history. Older cohorts with individuals with high age at migration may be more common among certain ethnic groups.

Westerners and Eastern and Southern Europeans are also to a lower degree overeducated and, conversely, to a higher degree undereducated if they have experienced a longer stay in Sweden. However this is not the case for non-Western immigrants. They have a divergent profile for undereducation which fluctuates around 20 % between the years since migration categories. Overeducation, on the other hand, follows the decreasing trend. The mean values of years since migration are about 21 for Westerners, and 16 years for Eastern and Southern Europeans and the non Western immigrants.

The tendency of overeducation to decrease with duration of residence seems to be clear. Note that this is not the case for undereducation; none of the ethnic groups seems to have a monotonous decrease in the rates of undereducation. These rates fluctuate more or less for all immigrant groups at least once. This could be linked to the weaker theoretical support for the effect of years since migration on the incidence of undereducation.

**Table 7 Years Since Migration and Mismatch Among Eastern and Southern European and Non- Western Immigrants (RM)**

Years Since Migration	Eastern & Southern Europé				Non-Western			
	Normal	Undereducated	Overeducated	Total	Normal	Undereducated	Overeducated	Total
6-10	11440	2200	5305	18945	5778	2667	3493	11938
	<b>60,4%</b>	<b>11.6%</b>	<b>28.0%</b>	<b>100.0%</b>	<b>48.4%</b>	<b>22.3%</b>	<b>29.3%</b>	<b>100.0%</b>
11-15	3101	736	1194	5031	12599	4856	5364	22819
	<b>61,6%</b>	<b>14.6%</b>	<b>23.7%</b>	<b>100.0%</b>	<b>55.2%</b>	<b>21.3%</b>	<b>23.5%</b>	<b>100.0%</b>
16-20	2217	586	614	3417	8312	2556	2309	13177
	<b>64,9%</b>	<b>17.1%</b>	<b>18.0%</b>	<b>100.0%</b>	<b>63.1%</b>	<b>19.4%</b>	<b>17.5%</b>	<b>100.0%</b>
21-25	2058	581	545	3184	5959	1852	1295	9106
	<b>64,6%</b>	<b>18.2%</b>	<b>17.1%</b>	<b>100.0%</b>	<b>65.4%</b>	<b>20.3%</b>	<b>14.2%</b>	<b>100.0%</b>
26-30	1534	619	264	2417	4320	1243	990	6553
	<b>63,5%</b>	<b>25.6%</b>	<b>10.9%</b>	<b>100.0%</b>	<b>65.9%</b>	<b>19.0%</b>	<b>15.1%</b>	<b>100.0%</b>
31+	3672	1148	541	5361	1137	330	239	1706
	<b>68,5%</b>	<b>21.4%</b>	<b>10.1%</b>	<b>100.0%</b>	<b>66.6%</b>	<b>19.3%</b>	<b>14.0%</b>	<b>100.0%</b>
<b>Total</b>	24022	5870	8463	38355	38105	13504	13690	65299
	<b>62,6%</b>	<b>15.3%</b>	<b>22.1%</b>	<b>100.0%</b>	<b>58.4%</b>	<b>20.7%</b>	<b>21.0%</b>	<b>100.0%</b>

The JA measure results in similar patterns (see table 8 and 9). The major overall difference in comparison to the RM based results is the much lower percentage correctly matched and the higher shares of undereducation. Nordic immigrants are now the ones with a divergent profile of undereducation which increases moderately from a high initial value.

**Table 8 Years Since Migration and Mismatch for Nordic and Western Immigrants (JA)**

Years Since Migration	Nordic				Western			
	Normal	Undereducated	Overeducated	Total	Normal	Undereducated	Overeducated	Total
6-10	860	578	273	1711	1302	680	539	2521
	<b>50.3%</b>	<b>33.8%</b>	<b>16.0%</b>	<b>100.0%</b>	<b>51.6%</b>	<b>27.0%</b>	<b>21.4%</b>	<b>100.0%</b>
11-15	1769	1160	388	3317	854	668	334	1856
	<b>53.3%</b>	<b>35.0%</b>	<b>11.7%</b>	<b>100.0%</b>	<b>46.0%</b>	<b>36.0%</b>	<b>18.0%</b>	<b>100.0%</b>
16-20	1227	839	232	2298	651	530	196	1377
	<b>53.4%</b>	<b>36.5%</b>	<b>10.1%</b>	<b>100.0%</b>	<b>47.3%</b>	<b>38.5%</b>	<b>14.2%</b>	<b>100.0%</b>
21-25	1954	1370	384	3708	620	514	169	1303
	<b>52.7%</b>	<b>36.9%</b>	<b>10.4%</b>	<b>100.0%</b>	<b>47.6%</b>	<b>39.4%</b>	<b>13.0%</b>	<b>100.0%</b>
26-30	3315	2467	584	6366	707	560	184	1451
	<b>52.1%</b>	<b>38.8%</b>	<b>9.2%</b>	<b>100.0%</b>	<b>48.7%</b>	<b>38.6%</b>	<b>12.7%</b>	<b>100.0%</b>
31+	9290	6757	1409	17456	1204	955	243	2402
	<b>53.2%</b>	<b>38.7%</b>	<b>8.1%</b>	<b>100.0%</b>	<b>50.1%</b>	<b>39.8%</b>	<b>10.1%</b>	<b>100.0%</b>
<b>Total</b>	18415	13171	3270	34856	5338	3907	1665	10910
	<b>52.8%</b>	<b>37.8%</b>	<b>9.4%</b>	<b>100.0%</b>	<b>48.9%</b>	<b>35.8%</b>	<b>15.3%</b>	<b>100.0%</b>

**Table 9 Years Since Migration and Mismatch Among Eastern and Southern European and Non- Western Immigrants (JA)**

Years Since Migration	Eastern & Southern Europe				Non-Western			
	Normal	Undereducated	Overeducated	Total	Normal	Undereducated	Overeducated	Total
6-10	11173	2503	5269	18945	5413	2335	4190	11938
	<b>59.0%</b>	<b>13.2%</b>	<b>27.8%</b>	<b>100.0%</b>	<b>45.3%</b>	<b>19.6%</b>	<b>35.1%</b>	<b>100.0%</b>
11-15	2808	1008	1215	5031	11360	5069	6390	22819
	<b>55.8%</b>	<b>20.0%</b>	<b>24.2%</b>	<b>100.0%</b>	<b>49.8%</b>	<b>22.2%</b>	<b>28.0%</b>	<b>100.0%</b>
16-20	1882	876	659	3417	7206	3299	2672	13177
	<b>55.1%</b>	<b>25.6%</b>	<b>19.3%</b>	<b>100.0%</b>	<b>54.7%</b>	<b>25.0%</b>	<b>20.3%</b>	<b>100.0%</b>
21-25	1724	890	570	3184	4810	2576	1720	9106
	<b>54.1%</b>	<b>28.0%</b>	<b>17.9%</b>	<b>100.0%</b>	<b>52.8%</b>	<b>28.3%</b>	<b>18.9%</b>	<b>100.0%</b>
26-30	1245	851	321	2417	3371	2103	1079	6553
	<b>51.5%</b>	<b>35.2%</b>	<b>13.3%</b>	<b>100.0%</b>	<b>51.4%</b>	<b>32.1%</b>	<b>16.5%</b>	<b>100.0%</b>
31+	2962	1780	619	5361	835	631	240	1706
	<b>55.3%</b>	<b>33.2%</b>	<b>11.5%</b>	<b>100.0%</b>	<b>48.9%</b>	<b>37.0%</b>	<b>14.1%</b>	<b>100.0%</b>
<b>Total</b>	21794	7908	8653	38355	32995	16013	16291	65299
	<b>56.8%</b>	<b>20.6%</b>	<b>22.6%</b>	<b>100.0%</b>	<b>50.5%</b>	<b>24.5%</b>	<b>24.9%</b>	<b>100.0%</b>

These tables conclude the study of the incidence of labour market mismatch. The measurements give similar results when studying the outcome for various ethnic groups differing in years since migration and the location where they mainly were brought up.

Similar patterns can be found by studying how the incidence of labour market mismatch varies with potential experience. This is expected as the two variables are closely related and both considered important for labour market outcomes by theory. The regression analysis in the next section will attend to the issue of which of these two variables is most important for the mismatch outcome.

According to these results and with the achievement interpretation of the mismatch, which is in line with the findings of ORU studies and the modified *Search and Match Theory*, there is a case for the statement that non-Western immigrants are to greater extent overachievers compared to Western ones. This is the case if undereducation is linked to a higher return relative to what would be the case if the immigrant had a job with a requirements that fitted the her attained education. Moreover, the modified *Search and Match Theory* states that individuals only change jobs if they are overeducated to attain higher levels of wealth. This result is not the conclusion we draw if we problemize the issue as in the case of age of migration above by combining components of RM and JA. This is done in table 10:

Table 10 Labour Outcome and Requirements Among Ethnic Groups

Labour Market Outcome (RM)			Requirement (JA)				Total
			No Educational Requirements	Secondary	Lower Tertiary	Higher Tertiary	
Normal		n	1464	13552	3618	3025	21659
	Nordic Countries	%	6.8%	62.6%	16.7%	14.0%	100.0%
		n	360	2469	1308	2436	6573
	Western	%	5.5%	37.6%	19.9%	37.1%	100.0%
	Eastern and Southern Europe	n	2708	15820	2234	3260	24022
		%	11.3%	65.9%	9.3%	13.6%	100.0%
	Non- Western	n	5661	21866	3859	6719	38105
		%	14.9%	57.4%	10.1%	17.6%	100.0%
	Total	n	10193	53707	11019	15440	90359
		%	11.3%	59.4%	12.2%	17.1%	100.0%
Undereducated		n	899	7372	784	775	9830
	Nordic Countries	%	9.1%	75.0%	8.0%	7.9%	100.0%
		n	158	946	253	506	1863
	Western	%	<b>8.5%</b>	<b>50.8%</b>	<b>13.6%</b>	<b>27.2%</b>	<b>100.0%</b>
	Eastern and Southern Europe	n	1007	4083	281	499	5870
		%	17.2%	69.6%	4.8%	8.5%	100.0%
	Non- Western	n	3160	8509	664	1171	13504
		%	<b>23.4%</b>	<b>63.0%</b>	<b>4.9%</b>	<b>8.7%</b>	<b>100.0%</b>
	Total	n	5224	20910	1982	2951	31067
		%	16.8%	67.3%	6.4%	9.5%	100.0%
Overeducated		n	233	1962	682	490	3367
	Nordic Countries	%	6.9%	58.3%	20.3%	14.6%	100.0%
		n	165	1024	683	602	2474
	Western	%	<b>6.7%</b>	<b>41.4%</b>	<b>27.6%</b>	<b>24.3%</b>	100.0%
	Eastern and Southern Europe	n	1052	5857	991	563	8463
		%	12.4%	69.2%	11.7%	6.7%	100.0%
	Non-Western	n	2044	8479	1900	1267	13690
		%	<b>14.9%</b>	<b>61.9%</b>	<b>13.9%</b>	<b>9.3%</b>	100.0%
	Total	n	3494	17322	4256	2922	27994
		%	12.5%	61.9%	15.2%	10.4%	100.0%

The relevant distributions for this discussion are marked in bold. It is clearly seen that Western European immigrants are to a much greater extent undereducated in occupations with lower and higher tertiary requirements (about 41%) compared to non Western immigrants (about 14 %). This relationship is also existent in the case of overeducation.

## 6.4 The Magnitude of Mismatch

Table 11 presents the mean years of overeducation among the different ethnic groups which gives a rough picture of which group has most years of over- and undereducation. This is done by using the ORU approach based on the RM measure, described in section 3.2.

**Table 11 Mean Years of Under- and Overeducation Among The Ethnic Groups**

Origin	Years of Undereducation		Years of Overeducation	
	Mean		Mean	
Nordic Countries	Mean	<b>.9852</b>	<b>1.0349</b>	
	n	9830	3367	
	Std. Deviation	.62480	1.00236	
Western	Mean	<b>.9646</b>	<b>1.4113</b>	
	n	1863	2474	
	Std. Deviation	.80231	1.24501	
Eastern and Southern Europe	Mean	<b>.9663</b>	<b>1.4549</b>	
	n	5870	8463	
	Std. Deviation	.65675	1.19386	
Non-Western	Mean	<b>.9027</b>	<b>1.5777</b>	
	n	13504	13690	
	Std. Deviation	.67015	1.22373	
Total	Mean	<b>.9445</b>	<b>1.4606</b>	
	n	31067	27994	
	Std. Deviation	.66346	1.20401	

The Nordic immigrants seem to have the highest mean years of undereducation but the lowest mean years of overeducation. Immigrants from countries outside Europe, on the other hand, have the lowest mean years of undereducation and the highest mean years of overeducation.

This table seems to hint that immigrants from Nordic countries have the most favorable labour market outcome if we in general associate or identify undereducation with overachievement and overeducation with underachievement. The table is ordered in descending order with respect to achievement in the labour market. Immigrants from non western countries seem to be the ones with the least favorable labour market outcome in this regard.

## 6.5 Summary

A surprisingly strong case for undereducation increasing with years of residence can be made by the results. The modified search and match theory is hence supported by this preliminary evidence. In spite of this, the rates of undereducation fluctuate between many categories of years since migration so the increase is not monotonous. This could be a reflection of the

weaker theoretical support of the effect of duration of residence on the incidence of undereducation. There was a stronger theoretical case for the decrease of overeducation with years since migration which is also supported by the data. These results are valid using both the RM and JA methods. The age at migration seems to be an important determinant of the mismatch outcome. This fact was expected because having attended Swedish compulsory school and being brought up in Sweden are expected to alter the transferability of human capital. Moreover a low age at migration is related to a high number of years since migration in the sample. This issue is approached with regression analysis in section seven.

Cohort effects are thought to affect the figures. The ethnic groups differ in composition, and older cohorts with individuals with high age at migration may be more common among certain groups.

A ‘blind’ spot of the RM method was analyzed by combining it with the required education measure which is used in the construction of the JA measure of mismatch. This analysis revealed that immigrants from non-Western countries are to a greater extent over- and undereducated in occupations with low requirements. These occupations are arguably associated with lower status and income.

The analysis of the data hints that Nordic and Western immigrants are the ones with the best labour market situation and that non-Western immigrants have the least favorable situation in terms of labour market mismatch.

## **7 Regression Analysis of the Mismatch among Immigrants**

### **7.1 Specification**

The reason for this study is to take different explanatory variables suggested by theory in to account when analyzing the differences in the incidence of mismatch among immigrant groups. By doing this, the differences in the probability between two groups of immigrants can be computed, given the set of explanatory variables. That is e.g., the differences in the probability of being overeducated which are not explained by *years since migration* and *potential experience*. If all the relevant explanatory variables are included, then the estimated difference can be associated with the ethnicity of the immigrant. This part of the analysis is therefore an important part of the aim of the thesis, namely to study differences in the incidence of overeducation between various ethnic groups.

The transferability of human capital could in theory be one of the most important determinants of the mismatch outcomes. A low age at migration can be expected to influence the transferability of human capital in a positive way. This should in turn imply that immigrants with a low age at migration are overeducated to a smaller extent than the ones with a high age at migration.

This hypothesis is tested by dividing each immigrant group in to subgroups according to whether their age at migration is above or under 16. The probability of being overeducated is hypothesized to be lower among immigrants who arrived to Sweden at a low age, controlling for civil status and citizenship, *potential experience* and *years since migration*.

The analysis is carried out with a linear probability model (LPM). This method has been criticized by econometricians but is common in applied work. This simpler tool serves the purpose of this study well as it will be mainly used to make simple distinctions in the probability of being under and overeducated between relevant groups and at the same time controlling for other relevant explanatory variables. The motivation to use this model is that in this case, it could be like “killing a fly with cannon” when using a more complex multinomial model, which is expected to generate similar results anyway.

The analysis begins with estimations of the probability of being over- and undereducated among all immigrant men in the data set while controlling for partner, citizen status, potential experience and years since migration. These results are reported in section 7.2.1

The second set of models distinguishes between four immigrant groups; Nordic (base group), Western, Eastern and Southern Europeans and non-Western immigrants. These results are reported in section 7.2.2

Nordin (2007) stated that the set of immigrants could be divided in two groups in respect of the importance or significance of the duration of residence and experience. The groups are on one hand Western and Nordic Europeans and on the other hand Eastern & Southern Europeans and non European immigrants. Western European and Nordic immigrants could have a shorter assimilation time and higher transferability of human capital. If so, years since migration could have a lower economic significance and experience a higher one. The opposite could be true for the other immigrant groups. This difference is supposedly connected to the different profiles of adaptation time and in the transferability of human

capital. (Nordin 2007, p.56) This suggests a model which allows for different slopes in respect to *ysm* and *exp*.

The third step in the analysis is to distinguish the four groups by dividing the immigrants after age at migration which gives eight distinct groups of immigrants. These results are reported in section 7.2.3

Two types of models are estimated, one for overeducation and one for undereducation. The difference between these models is the dependent variable. In the overeducation model the binary variable equals one if overeducated and zero otherwise. On the contrary, the variable equals one if undereducated and zero otherwise for the undereducation estimations. The dependent variable is based on the RM measure.

The statistical equation is:

$$y = \beta_0 + \delta_0 \text{Citizen} + \delta_1 \text{Partner} + \delta_2 \text{NU16} + \delta_3 \text{WU16} + \delta_4 \text{WO16} + \delta_5 \text{E\&SU16} \\ + \delta_6 \text{E\&SO16} + \delta_7 \text{OEU16} + \delta_8 \text{OEO16} + \beta_1 \text{YSM} + \beta_2 \text{YSM}^2 + \beta_3 \text{EXP} \\ + \beta_4 \text{EXP}^2$$

The binary immigrant group variables can take a value of zero or one. E.g. **WU16** stands for Western European immigrants who migrated at an age under 16, **WO16** is for the set of Western immigrants who migrated at an age over 16. There are two similar binary variables for each set of immigrants partitioned by the age of migration. The only exception is Nordic immigrants who migrated at an age over 16. This set of immigrants is the base group. From this specification it is possible to analyze the differences in the probability of being overeducated respectively undereducated between groups.

## 7.2 Results

### 7.2.1 Overeducation and Undereducation among All Immigrants

The probabilities of being over- or undereducated among all immigrants of the two initial models for labour market mismatch seem to be in line with the overall trends outlined in the

earlier descriptive analysis. The quadratic term in the overeducation case was omitted (negative sign) because it was not statistically significant at the 5% level.

The Overeducation model suggests that *potential experience* lowers the probability of being overeducated at all levels *potential experience*. The variable *years since migration* lower on the other hand the probability of being overeducated at a decreasing rate in the approximately first 39 years (44 due to the restriction of immigrants with at least 5 years since migration).

**Table 12 Labour Market Mismatch among all Immigrants**

	Overeducation	Undereducation
(Constant)	.473*** (.000)	.226*** (.000)
sv_medb	.031*** (.000)	-.052*** .000
partner	.033*** .000	-.020*** (.000)
ysm	-.0156*** (.000)	.00327*** (.000)
ysmsqr	.0002*** .000	-.00009*** (.000)
exp	-.006*** (.000)	-.016*** .000
expsqr	-	.001*** (.000)
R <sup>2</sup>	.053	.090
N	149420	149420

In the case of undereducation, *potential experience* seems to lower the probability of being undereducated at a decreasing rate initially, i.e. the first eight years of experience. After these eight initial years, the probability of undereducation increases with *potential experience*. On the contrary, *years since migration* increases the probability of undereducation in the first 18 year and lowers the probability of undereducation (marginally) thereafter.

The estimations suggest that the processes which determine the undereducation- and overeducation outcomes are almost the opposite in terms of the effects of *years since migration* and *potential experience*, but with additional differences.

Potential experience has an unambiguous effect in the case of overeducation, but not in the case of undereducation. This is arguably also true for *years since migration*. The duration of residence in Sweden has a decreasing negative effect for about four decades on the probability

of overeducation. There seems to be an adaptation phase during the two first decades when the duration of residence increases the probability of undereducation, *years since migration* lowers the probability (slightly) thereafter. This suggests e.g. that an experienced immigrant improves her chances of being undereducated and is less likely to experience overeducation with each year since migration for the first two decades in Sweden. Thereafter the probability of undereducation decreases marginally with duration of residence while the effect on the probability of overeducation is unchanged.

### **7.2.2 Labour Market Mismatch among Four Ethnic Groups**

The estimated models in this section make a distinction between Nordic, Western, Eastern and Southern Europeans and non-Western immigrants. From the analysis in the previous section it seemed like *years since migration* and *potential experience* both lower the probability of being overeducated; the latter was best captured as a constant effect. The quadratic terms for *years since migration* are omitted in this version because these terms seem to cause problems with the statistical significance of the estimated coefficients associated with years since migration. The results from the estimations are presented in table 13.

The undereducation model has quadratic terms associated with *years since migration* and *potential experience* but seems to fit the data best assuming a common slope for the different groups.

Westerners are initially the most likely of being overeducated, when controlling for the other factors followed by Non-Western immigrants. The estimated coefficients for *years since migration* and *potential experience* have an interesting interpretation. The negative effect of *years since migration* is greatest for Non-Western immigrants (-.007), followed by Eastern and Southern Europeans (-.006), Western and Nordic Immigrants have the smallest estimated effects in this regard (-.004 and -.002 respectively). The inverse relation is true in the case of the negative effect of *potential experience*. In this case Western Europeans are associated with the greatest negative effect (-.01) followed by Nordic (-.007), non-Western (-.006) and Eastern and Southern European immigrants (-.005) in descending order. This means that years since migration is more important for non-Western and Eastern and Southern European immigrants than *potential experience* for the probability of becoming overeducated. The converse is true for Nordic and Western immigrants for whose labour market experience have the greatest impact in this regard.

**Table 13 Labour Market Mismatch among Four Immigrant Groups**

Variables	Overeducation	Undereducation
(Constant)	.301*** (.008)	.157*** (.013)
Swedish Citizen	.028*** (.003)	-.048*** (.003)
Partner	.032*** (.002)	-.014*** (.002)
Western	.199*** (.013)	.022 (.018)
Eastern and Southern	.065*** (.010)	.036* (.015)
Non-Western	.090*** (.009)	.219*** (.015)
YSM	-.002*** (.000)	0.0115*** (.001)
WesternxYSM	-.002*** (.000)	-0.0078*** (.002)
Eastern and SouthernxYSM	-.004*** (.000)	-0.0100*** (.001)
Non-WesternxYSM	-.005*** (.000)	-0.0192*** (.001)
YSM <sup>2</sup>	-	-0.0002*** (.000)
WesternxYSM <sup>2</sup>	-	0.0001*** (.000)
Eastern and SouthernxYSM <sup>2</sup>	-	0.0002*** (.000)
Non-WesternxYSM <sup>2</sup>	-	0.0003*** (.000)
EXP	-.007*** (.000)	-.017*** (.001)
WesternxEXP	-.003*** (.001)	-
Eastern and SouthernxEXP	.002*** (.000)	-
Non-WesternxEXP	.001*** (.000)	-
EXP <sup>2</sup>		.001*** (.000)

These patterns are not found in the case of undereducation. Instead, *years since migration* matters more for Nordic and Western immigrants, it lowers in fact the probability of undereducation for non-Western immigrants but with a small amount and at a decreasing rate. Experience is assumed to have the same effect for all groups and it behaves almost exactly as in the previous set of models without distinction between ethnic groups.

These findings are to some extent consistent with the weaker theoretical case for effect of *years since migration* and the ambiguous results from the analysis with descriptive statistics, especially in the case of non-Western immigrants. The overeducation model supports the hypothesis of differences in the transferability of human capital between the group consisting

of Nordic and Western immigrants on one hand, and non-Western and Eastern and Southern Europeans on the other.

### 7.2.3 Models with Ethnicity and Age at Migration

#### 7.2.3.1 Overeducation

The contribution of *years since migration* to the probability of being overeducated is given by  $-.010 + 2 * 0.023ysm > 0$  for all years since migration (see table 14). The contribution of years since migration on the probability of being overeducated seems to be erroneously high at the mean value of *ysm*. For experience it is  $-0.013+2*0.007exp = 0$  so there is a positive impact of experience within a year. This suggests that the model captures the contributions of the continuous variables poorly. Theory and descriptive statistics suggest that years since migration and experience should lower the probability of overeducation. Therefore the model is reduced by dropping the quadratic forms. This in turn makes the contribution of *ysm* statistically and economically insignificant so a model without this variable is computed (see column 2 in table 14), this specification keeps the qualitative features of the original model and  $R^2$  is not severely affected.

**Table 14 LPM for Overeducation**

	(1)	(2)
(Constant)	.496 (.008)	.360 (.005)
Swe Citizen	.039 (.003)	.037 (.003)
Partner	.025 (.002)	.028 (.002)
Nordic Under 16	-.112 (.005)	-.083 (.004)
West Under 16	-.071 (.008)	-.039 (.007)
West Over 16	.088 (.005)	.096 (.005)
East & South Under 16	-.091 (.006)	-.081 (.006)
East & South Over 16	.041 (.004)	.056 (.004)
Non-Western Under16	-.135 (.005)	-.135 (.005)
Non-Western Over16	.038 (.004)	.043 (.004)
YSM	-.010 (.000)	-
YSM <sup>2</sup>	.023 (.001)	-
EXP	-.013 (.001)	-.010 (.000)
EXP <sup>2</sup>	.007 (.001)	-
R <sup>2</sup>	.064	.06
N	149420	149420

This statistical specification is easier to interpret but it can not be used to test whether the differences are statistically significant or not. In the interaction formulation the ethnicity and age at migration appear separately (See Appendix, table A1). With this formulation it is possible to test the null hypothesis which states that the difference in the probability of being overeducated among ethnic groups does not depend on age at migration. This equivalent statistical specification with interaction terms is used for this purpose. (see Woolridge 2006, p.244) This regression shows that all differences are significant at the 1 % level for both

formulations above i.e. with and without quadratic forms and *ysm*.<sup>11</sup> This means that a partly Swedish compulsory schooling or upbringing is a significant explanatory variable for the incidence of overeducation. Each year of experience lowers on average the probability of overeducation by one percent, a feature which is theory consistent.

By setting all groups except the Nordic group to zero, the difference in the probability of being overeducated between Nordic immigrants with a low and high age at migration (under 16 and over 16) is calculated. This difference is  $\delta_2 = (-0.112)$ , i.e. Nordic immigrants who migrated at an age under 16 are about 12 % less likely of being overeducated than those who migrated at an age over 16. In the same way, the difference between non- Western immigrants with a low age of migration and Nordic immigrants with a high age of migration is  $\delta_7 = (-.135)$ , i.e. immigrants from non- Western countries with compulsory schooling which is partly Swedish are less likely to be overeducated than their Nordic counterparts with foreign compulsory schooling. On the other hand, when the foreign education situation for non- Westerners is the opposite (i.e. arrived when age at migration > 16 and Nordic when age at migration < 16), then the group of immigrants from countries outside of Europe are more likely to experience overeducation ( $\delta_8 = .038$ ) than the base group.

This relationship is a recurrent for all ethnic groups, i.e. they are more likely to be overeducated relative to the Nordic base group if the age at migration was high and conversely less likely if they immigrated before the age of 16.

The tables 15 and 16 below can be seen as tables in which the elements represent the estimated differentials of the probability of being overeducated among ethnic groups who differ in their age at migration. The first element is the difference between the same group of Nordic immigrants and is therefore zero; the second element in the first row is the difference between Nordic immigrant with a low age of migration and Nordic immigrant with an age of migration over 16 and so on. The differentials above and below the diagonal are of the same magnitude, but with opposite signs.

From the tables 15 and 16 it is possible to get an overview of the impact of the immigrant's compulsory schooling /place of upbringing. The differentials *within* ethnic subgroups with different age at migration (i.e. over and below 16) are in general much greater than the

---

<sup>11</sup> These estimations are presented in table A1 of the appendix. By setting the right combinations of zeros and one the same coefficient estimates can be acquired but may be with marginal rounding errors.

differences *between* ethnic groups who share the same age at migration migration (e.g. Nordic and non- Western both with an age of migration under 16). These differences are the colored ones in table 15. As stated above, a low migration age is linked to a lower probability of overeducation relative to a high age at migration.

**Table 15 Probability Differentials for Overeducation**

	Nordic under16	Nordic Over 16	west under 16	west over 16	East &South under 16	East&South over 16	Outside Europe under 16	Outside Europe over 16
Nordic under16	0.0%	-11.2%	-4.1%	-20.0%	-2.1%	-15.3%	2.3%	-15.0%
Nordic Over 16	11.2%	0.0%	7.1%	-8.8%	9.1%	-4.1%	13.5%	-3.8%
west under 16	4.1%	-7.1%	0.0%	-15.9%	2.0%	-11.2%	6.4%	-10.9%
west over 16	20.0%	8.8%	15.9%	0.0%	17.9%	4.7%	22.3%	5.0%
East&South under 16	2.1%	-9.1%	-2.0%	-17.9%	0.0%	-13.2%	4.4%	-12.9%
East&South over 16	15.3%	4.1%	11.2%	-4.7%	13.2%	0.0%	17.6%	0.3%
OutsideEurope under 16	-2.3%	-13.5%	-6.4%	-22.3%	-4.4%	-17.6%	0.0%	-17.3%
Outside Europe over 16	15.0%	3.8%	10.9%	-5.0%	12.9%	-0.3%	17,3%	0.0%

This feature is directly presented in table 16 which contains the differentials *within* ethnic groups in the diagonal.

**Table 16 Age of Migration Differentials**

	Nordic Over 16	west over 16	East&South over 16	Outside Europe over 16
Nordic under 16	-11.2%	-20.0%	-15.3%	-15.0%
west under16	-7.1%	-15.9%	-11.2%	-10.9%
East&South under16	-9.1%	-17.9%	-13.2%	-12.9%
OutsideEurope under16	-13.5%	-22.3%	-17.6%	-17.3%

The diagonal of table 16 contains differentials which are overall much greater than the differences between ethnic groups with similar profile in age of migration (colored elements in table 15). This means that age at migration is an important determinant for the incidence of overeducation. The probability of being overeducated is to a greater extent influenced by being brought up in Sweden and Swedish compulsory schooling rather than belonging to one of the four ethnic groups. If two immigrants with different age at migration (one over 16 and the other under 16) but from the same ethnic group are compared, then the difference in the probability of being overeducated between e.g. two non-Western immigrants is on average greater than between a non-Westerner and a Western immigrant if they both arrived to Sweden at a low age (under 16).

### 7.2.3.2 Undereducation

A similar regression is carried out with undereducation as the incidence of success with the same type of table for the differences between groups.<sup>12</sup>

$R^2$  is now slightly higher and as expected, the pattern from the overeducation regression is now the reversed (see table 17). Immigrant groups with Swedish education are all more likely to be undereducated than the Nordic group with foreign education or upbringing. Conversely, immigrant groups with foreign education are less likely to be undereducated relative their Nordic counterparts.

<sup>12</sup> The significance of the coefficients is presented in table A2 in the Appendix.

Years since migration seems confusingly enough to lower the probability of undereducation slightly, experience on the other hand seems to improve the probability of undereducation which is a theory consistent feature. A reduced model without the quadratic forms does not improve the model in terms of interpretation. This could be interpreted to mean that years since migration is less important than experience for the probability of being undereducated. The ambiguous result could in fact be in line with the 'weaker' theoretical support for undereducation to decline with the duration of residence in Sweden.

**Table 17 LPM for Undereducation**

	(3)	(4)
(Constant)	.246 (.008)	.027 (.005)
sv_medb	-.055 (.003)	-.055 (.003)
partner	-.009 (.002)	-.019 (.002)
Nordic Under16	.135 (.006)	.087 (.005)
West Under16	.100 (.008)	.073 (.008)
West Over16	-.103 (.005)	-.122 (.005)
East and South under16	.079 (.006)	.062 (.006)
East and South over16	-.108 (.004)	-.127 (.004)
Non-Western under16	.131 (.006)	.155 (.006)
Non-Western over16	-.025 (.004)	-.052 (.004)
Ysm	-.004 (.000)	-.007 (.000)
YSM <sup>2</sup>	-.008 (.001)	-
EXP	-.011 (.001)	.018 (.000)
EXP <sup>2</sup>	.064 (.001)	-
R <sup>2</sup>	.108	.094
N	149420	149420

The significance of the difference-coefficients is tested with an interaction specification of the model. This regression is presented in the appendix (table A2) and shows that all interaction terms are statistically significant at the one percent level.

The patterns of the probability differentials are almost the opposite of the overeducation case. Differences among immigrant groups with Swedish and foreign education tend to be

larger than differences between groups of immigrants with the same profile of compulsory education. E.g. the difference between an immigrant from a Nordic country with a Swedish compulsory education and a foreign compulsory education is on average greater than the difference between a Nordic and non-European immigrant if they both have Swedish compulsory schooling. The difference between the latter is in fact almost negligible. Each year since migration seems to lower the probability of being undereducated oddly enough.

These patterns are illustrated in the difference tables below (table 18 and 19). Coloured elements denote the differences between ethnic groups of the same education profile (i.e. within the sets of Swedish or foreign education). These differences are in general smaller than the ones within ethnic groups with differences in compulsory schooling.

**Table 18 Probability Differentials for Undereducation**

	Nordic under 16	Nordic Over 16	West under 16	West over 16	East & South under 16	East & South over 16	Non-Western under 16	Non-Western over 16
Nordic under 16	0.0%	13.5%	3.5%	23.8%	5.6%	24.3%	0.4%	16,0%
Nordic Over 16	-13.5%	0.0%	-10.0%	10.3%	-7.9%	10.8%	-13.1%	2.5%
west under 16	-3.5%	10.0%	0.0%	20.3%	2.1%	20.8%	-3.1%	12.5%
west over 16	-23.8%	-10.3%	-20.3%	0.0%	-18.2%	0.5%	-23.4%	-7.8%
East & South under 16	-5.6%	7.9%	-2.1%	18.2%	0.0%	18.7%	-5.2%	10.4%
East & South over 16	-24.3%	-10.8%	-20.8%	-0.5%	-18.7%	0.0%	-23.9%	-8.3%
Non-Western under 16	-0.4%	13.1%	3.1%	23.4%	5.2%	23.9%	0.0%	15.6%
Non-Western over 16	-16.0%	-2.5%	-12.5%	7.8%	-10.4%	8.3%	-15.6%	0.0%

By comparing Table 18 with the colored elements of table 19 it can be clearly seen that the differentials within a particular ethnic group with a different age at migration (i.e. over and

under 16) are in general greater than any difference between ethnic groups of the same migration age.

**Table 19 Age at Migration Differentials**

	Nordic Over 16	west over 16	East & South over 16	Outside Europe over 16
Nordic under 16	13.5%	23.8%	24.3%	16.0%
west under 16	10.0%	20.3%	20.8%	12.5%
East & South under 16	7.9%	18.2%	18.7%	10.4%
Outside Europe under 16	13.1%	23.4%	23.9%	15.6%

### 7.3 Quandaries

It is difficult to find the exact problems with the regression regarding the counterintuitive sign of the *years since migration* when controlling for age at migration. There are some possible explanations for this. In addition, the regressions are subject to potential problems of more general character.

Selection bias is an important issue. Clearly some unobserved feature may influence age at migration. Moreover, individuals with a low age at migration tend to have more years in Sweden than individuals with high age at migration. If *years since migration* lower the probability of overeducation and rises the probability of undereducation, then the categories of age at migration will be mere reflections of this. But if the variable of being brought up in Sweden or having Swedish compulsory schooling is omitted, then this will affect the estimate of *ysm*. In other words, age at migration is positively correlated with overeducation and negatively correlated with *years since migration*. (see Kennedy 2003, p.397f)

*Potential experience* and age are closely related but the latter should have the opposite sign of the former in most specifications. A specification that incorporates this information e.g. by using the ratio of these two is overlooked in this thesis. This omitted variable could alter the results somewhat if used. Other omitted variables of importance could obviously also affect the estimated coefficients, unobserved ability and controls for discrimination are two examples mentioned in the theory section.

It is possible that the models which controlled for age at migration lack enough variation to give correct estimates of the *ysm* coefficient. *Years since migration*, *age at migration* and *potential experience* are related in a way that may prevent accurate estimates, so it might be impossible to separate the effects of these three.

The variables *ysm* and *exp* are related in a problematic way. E.g., it turns out that *ysm* is equal to *exp* exactly when the individual migrated to Sweden the year when she finished school, assuming the Mincer proxy. An influential share of the sample may be quite close to this which should make it harder to distinguish the variables' individual effects in the regression analysis and result in less reliable estimates of the coefficients.<sup>13</sup>

Older cohorts have higher values of *ysm* and *exp* and may differ from younger cohorts in important ways, e.g. due to structural changes in the economy. This is mentioned initially, older cohorts may have a lower overall level of schooling etc.. This in turn affects the probability of labour market mismatch and must be considered when interpreting the effects of *ysm* and *exp*

#### 7.4 Summary

The estimated coefficients of *years since migration* and *potential experience* affects the probability of overeducation negatively and undereducation positively for a wide range of the variables in the specifications without *age at migration*. The way *ysm* and *exp* affects the probabilities of mismatch differ in the case of over- and undereducation. These continuous variables seem to have a more straight forward linear influence on the probability of overeducation. On the other hand, these continuous variables seem to have quadratic features in the undereducation case, specially *ysm*. This is found to be a characteristic of the models without age at migration, from the simplest to the more complicated ones which control for ethnic groups.

These findings are quite consistent with the stronger support for the negative effect of *ysm* on overeducation. The importance of different quadratic terms on *ysm* for each of the ethnic groups may reflect the weaker theoretical support of its effect on undereducation and is fairly consistent with the descriptive analysis. The long term behavior, which can be thought to be captured by the quadratic terms associated with *ysm* (small in size but which dominate for

---

<sup>13</sup>  $Ysm = 2003 - \text{Year of Migration}$ ,  $Exp = 2003 - \text{Year of Birth} - 7$ ,  $\text{Age at Migration} = \text{Year of Migration} - \text{Year of Birth}$ . So  $Exp = Ysm \Leftrightarrow 2003 - \text{Year of Birth} - 7 = 2003 - \text{Year of Migration} \Leftrightarrow \text{School} + 7 + \text{Year of Birth} - \text{Year of Migration} = \text{School} + 7 - \text{age at migration} = 0 \Leftrightarrow \text{School} + 7 = \text{Age at Migration}$

large values of  $ysm$ ), hints that the immigrant groups have different transferability of human capital, and are affected differently by the duration of residence in Sweden.

The overeducation model which distinguishes four ethnic groups suggests that Eastern and Southern European and non-Western immigrants have a lower transferability of human capital. The undereducation models (without age at migration) give substantial support to the modified *search and match theory*. In addition this estimate suggests that searching and matching is more efficient among Nordic and Western immigrants compared to among Eastern and Southern Europeans and specially non-Western immigrants, in terms of the probability of undereducation. In fact, non-Western immigrants tend to be less likely of becoming undereducated and less likely of becoming overeducated with each year of migration, according to the regression analysis. This suggests that non-Western immigrants are more likely to become correctly matched with duration of residence, a feature which is to some extent consistent with the descriptive statistics.

The reliability of the results with age at migration is somewhat questionable because of the unexpected sign of  $ysm$ . On the other hand, the somewhat simpler models which have the expected sign, suggest that the relationship between  $ysm$  and mismatch is intricate, at least in the case of undereducation. Moreover, both theory and the descriptive analysis indicated a weaker support for the effect of  $ysm$  on the probability of undereducation. Other possible explanations of the causes of the sign of  $ysm$  have been presented in this section.

The results from the models which control for age at migration hints that being brought up in Sweden and attending Swedish compulsory school, strongly affects the probability of mismatch. Immigrants with Swedish compulsory schooling are less likely to be overeducated and more likely to be undereducated than the base group. The results suggest that the probability of being over- or undereducated is to a greater extent determined by where the immigrant was raised or attended compulsory schooling than by ethnicity. These results suggest that the age at migration is more important than ethnicity in terms of the effect on the transferability of human capital.

## **8 Summary and Conclusions**

The analysis with descriptive statistics indicates a strong relationship between years since migration and the decrease of the share of overeducated. Moreover, years since migration are associated with an increase of the share of undereducated when data is analyzed in this way.

The importance of the immigrant's age at migration to Sweden is also considered in this part and is found to be plausible but beyond the scope of a descriptive analysis. This in turn motivates a more rigorous study which carried out in section seven.

These statistics give at a first glance strong support for the modified search and match theory, which is the only one which predicts an increase in the share of undereducated with the duration of residence in Sweden. The other theories which predict the decline of overeducation with duration of residence are also consistent with the results.

The labour market mismatch can be linked to the notion of over- and underachievement in the labour market, among other interpretations. The JA measure was found to be closer to this interpretation than the RM measure, but both of these measures conceal relevant aspects of achievement and mismatch.

A 'blind spot' can be analyzed when components of the two measures are combined. This is done by relating the mismatch outcome to the occupational requirements. This analysis reveals that non-Western immigrants to a much greater extent are mismatched in low requirement occupations, especially compared to Western immigrants who have a lower share undereducated than non-Westerners. If the mismatch is identified with achievement, then wrong conclusions about the relative achievement position could be drawn without further reflection. Another complementary way of analyzing this issue is by the magnitude of the mismatch that is used in the ORU approach, which deals with the returns of over- and undereducation.

The hypothesis which states that immigrants from Nordic and Western countries have a higher transferability of human capital than Eastern and Southern European and non-Western immigrants, is supported by the regression analysis which made a distinction between four ethnic groups. Potential labour market experience has a greater effect on the probability of being overeducated than duration of residence or years since migration among Nordic and Western immigrants. The converse was found to be the case for the two latter ethnic groups.

Undereducation seems to be subject to a different process than overeducation. In this case, years since migration affected the probability of undereducation more among Nordic and Western immigrants than Eastern and Southern Europeans, and especially non-Western immigrants. The modified search and match theory was the only one which predicted an

increase in undereducation with years since migration. With this interpretation in mind, the results indicate that Nordic and Western immigrants are the most successful in the wealth maximizing process of the searching and matching hypothesis. Non-western immigrants seem on the other hand not to be more likely to attain such positions with more years since migration to Sweden. The search and match process for these immigrants seem to result in occupations which fit their requirements to a greater extent instead.

The relationship between years since migration and the incidence of mismatch is not found in the linear probability model when controlling for age at migration, ethnicity, potential experience, years since migration, partner and citizen status. Potential experience does on the other hand increase the probability of undereducation and decreases the probability of overeducation.

The results from the first part of the analysis regarding age at migration are in a less ambiguous way, carried over to the linear probability model in the case of age at migration. A low age of migration is associated with a lower probability of being overeducated and a higher probability of being undereducated. The importance of Swedish compulsory schooling is further accentuated by how the patterns of differences in the probability of mismatch look like. The difference in the probability of being mismatched between two immigrants is in general much higher if they differ in age at migration than if they differ in ethnicity.

The result illuminates the importance of the transferability of human capital which is consistent with the theoretical adaptation of labour market theories in an immigrant context as outlined by Chiswick. The main determinant of a lower probability of mismatch is a low age of migration, hence being at least partly brought up in Sweden and having a Swedish compulsory education, not ethnicity per se, according to this set of models.

One interpretation of the Swedish labour market based on the regression analysis with age at migration is that ethnicity as such is not an important issue for the labour market mismatch. The other side of the coin is that the considerable importance of age at migration could be thought to imply a large threshold for immigrants who migrated as adults and are presumably less integrated.

One problem with the results from the regressions could be that the model does not take into account how the mismatch is distributed over job requirements or the magnitude of the mismatch. In addition, other important individual characteristics such as ability are not taken

into account. Dealing with unobserved heterogeneity in the characteristics of workers and occupations is presumably important to improve the models.

Immigrants from non-Western countries are in relation to the other ethnic groups, to a greater extent both over- and undereducated in jobs with low requirements. Moreover they are subject to the greatest magnitude of overeducation and the smallest magnitude of undereducation. In addition, these immigrants are not more likely to obtain occupations in which they are undereducated. These findings possibly hint that ethnicity-based discrimination should not be ruled out as a factor, but these facts could also be explained by a presumably lower transferability of human capital among non-Western immigrants and other unobserved features.

## 8 Appendix

### A. 1 The Significance of Age at Migration and Ethnicity

Dependent: Overeducation	(1)	(2)
(Constant)	.496*** (.008)	.360*** (.005)
Swedish Citizen	.039*** (.003)	.037*** (.003)
Partner	.025*** (.002)	.028*** (.002)
Western	.088*** (.005)	.096*** (.005)
EasterandSouthern	.041*** (.004)	.056*** (.004)
Outside	.038*** (.004)	.043*** (.004)
Age of migration under 16	-.112*** (.005)	-.083*** (.004)
WesternxAgemig_under16	-.048*** (.009)	-.052*** (.009)
EasterandSouthernxAgemig_under16	-.020*** (.007)	-.055*** (.007)
OutsidexAgemig_under16	-.062*** (.006)	-.096*** (.006)
ysm	-.010*** (.000)	
ysmsqr	.023*** (.001)	
exp	-.013*** (.001)	-.010*** (.000)
expsqr	.007*** (.001)	
R <sup>2</sup>	.064	.060
N	149420	149420

**A. 2 The Significance of Age at Migration and Ethnicity**

Dependent: Undereducation	(1)	(2)
(Constant)	.246*** (.008)	.027*** (.005)
Swedish Citizen	-.055*** (.003)	-.055*** (.003)
Partner	-.009*** (.002)	-.019*** (.002)
Western	-.103*** (.005)	-.122*** (.005)
EasterandSouthern	-.108*** (.004)	-.127*** (.004)
Outside	-.025*** (.004)	-.052*** (.004)
Age of migration under 16	.135*** (.006)	.087*** (.005)
WesternxAgemig_under16	.068*** (.009)	.108*** (.009)
EasterandSouthernxAgemig_under16	.053*** (.007)	.101*** (.007)
OutsidexAgemig_under16	.022*** (.006)	.120*** (.006)
ysm	-.004*** (.000)	-.007*** (.000)
ysmsqr	-.008*** (.001)	
exp	-.011*** (.001)	.018*** (.000)
expsqr	.064*** (.001)	
R <sup>2</sup>	.108	.094
N	149420	149420

## References

- Bourdet, Yves and Persson, Inga 2008. "Overeducation Among Swedish Youth" *Lund Working Paper* 2008:8, Department of Economics, Lund.
- Chiswick, Barry R. and Miller, Paul W. 2007. "The International Transferability of Immigrants' Human Capital Skills" *IZA DP* No. 2670.
- Hartog, Joop 1997 "Over-education and Earnings: "Where Are We, Where Should We Go?" *Economics of Education Review* vol. 19, pp. 131-147.
- Johansson, Mats. and Katz, Katarina. (2007). "Wage differences between women and men in Sweden - the impact of skill mismatch", *IFAU Working Paper* 2007:13, Institute for Labour Market Policy Evaluation (IFAU), Uppsala.
- Jones, Charles I. 2002 *Introduction to Economic Growth*, Second Edition W.W. Norton & Company.
- McGuinness, Séamus 2006. "Overeducation in the Labour Market" *Journal of Economic Surveys* vol. 20, No. 3 pp.387-418.
- Nordin, Martin 2007. "Studies in Human Capital, Ability and Migration" *Lund Economic Studies* No. 141.
- Orcutt, Harriet, Regets, Duleep and Regets, Mark C. "Immigrants and Human-Capital Investment" *The American Economic Review*, Vol 89, No. 2 pp.186-191.
- Sloane, Peter J. 2002. "Much Ado About Nothing –What Does the Over-Education Litterature Really Tell Us?" *International Conference on Over-education in Europe: What Do We Know?* Berlin, 22-23 November 2002.
- Varian, Hal R. 1992 *Microeconomic Analysis* Third Edition. W.W. Norton & Company.
- Verhaest, Dieter and Omeij Eddy 2006. "Measuring the Incidence of Over- and Undereducation" *Quality and Quantity* 40: pp.783-803.
- Voon, Derby and Miller, Paul W. 2005 "Undereducation and Overeducation in the Australian Labour Market" *The Economic Record*. Vol 81. Special Issue. August 2005 pp.22-33.

Wooldridge, Jeffrey M. 2006. *Introductory Econometrics*. Third Edition. Thomson South-Western.

**Internet Source**

*Standard för svensk yrkesklassifering* [http://www.scb.se/templates/Standard\\_36492.asp](http://www.scb.se/templates/Standard_36492.asp)